DOCUMENT RESUME

ED 146 921 IR 005 392

TITLE	Guidelines for the Establishment and Development of Multilingual Thesauri.
INSTITUTION	United Nations Educational, Scientific, and Cultural Organization, Paris (France).
REPORT NO	SC.76/WS/102
PUB DATE	Dec 76
NOTE	51p.
EDRS PRICE	MF-$0.83 Plus Postage. HC Not Available from EDRS.
DESCRIPTORS	*Construction (Process); English; French; German; *Guidelines; Indexing; Information Services; *Multilingualism; *Standards; *Thesauri
IDENTIFIERS	*Multilingual Thesauri

ABSTRACT

These guidelines are intended for the growing number of information services which use, or plan to use, a multilingual thesaurus. The purpose is to describe the various tasks involved in constructing and developing multilingual thesauri, and where possible, to recommend standard practices. The scope is restricted to a consideration of the problems of multilingualism as they relate to conventional thesauri--those intended for use by human indexers and searchers, designed for coordinate indexing, and which avoid semantic factoring. There is emphasis on thesauri which distinguish between descriptors and non-descriptors. Guidelines are based on experience gained in the fields of science and technology and they concentrate on the world's major languages in these fields. Definitions, methods of use and thesaurus construction, basic problems of multilingual thesauri, presentation and arrangement, the process of constructing multilingual thesauri, and organizational aspects are considered. (Author/KP)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED146921

UNITED NATIONS EDUCATIONAL,
SCIENTIFIC AND CULTURAL ORGANIZATION

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Guidelines for the Establishment and Development of Multilingual Thesauri

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

Unesco

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

IR005 392

Photographic reproduction
of the original manuscript

SC.76/WS/102

Paris, December 1976

Table of contents

1	Introduction	3
1.1	Background	3
1.2	Purpose	4
1.3	Scope	4
1.4	Limitations	5
2	Definitions	5
2.1	Definitions of a multilingual thesaurus	5
2.2	Approaches to the establishment and development of multilingual thesauri	6
2.2.1	Methods of use	7
2.2.2	Methods of thesaurus construction	7
3	The basic problems of multilingual thesauri	8
3.1	Degrees of equivalence	8
3.1.1	Exact equivalences	9
3.1.2	Inexact equivalences	11
3.1.3	Non-equivalences	23
3.1.4	Establishing equivalences	23
3.1.5	The adjustment of term relationships	27
3.2	The elements of a multilingual thesaurus	29
3.2.1	Descriptors	29
3.2.2	Proper names	29
3.2.3	Abbreviations and acronyms	30
3.2.4	Noun form and member	31
3.2.5	Homographs and polysemes	31
3.2.6	Definition scope notes	33
3.3	Relationships	35
3.3.1	Hierarchical relations	35
3.3.2	Associative relations	36
4	Presentation and arrangement	37
4.1	The introduction to the thesaurus	37
4.2	The main part of the thesaurus	37
4.3	Auxiliary parts of the thesaurus	37
4.3.1	Alphabetical indexes	37
4.3.2	Systematic listings and graphic displays	38
4.4	The relationship between the main and auxiliary parts of the thesaurus	38

5	The process of constructing a multilingual thesaurus	39
5.1	Preliminary procedures	39
5.1.1	Notification of appropriate clearing houses	39
5.1.2	Avoiding duplication of effort	39
5.2	Establishing equivalence classes	39
5.2.1	Ab initio method of construction	40
5.2.2	Translation of an existing (source) thesaurus into a target thesaurus	41
5.2.3	Merging of existing thesauri	42
5.3	Control procedures	42
5.4	Testing	48
6	Organizational aspects	48

1 Introduction

1.1 Background

The trend towards internationalism in information systems is exemplified by the UNISIST programme and by projects like AGRIS and INIS. Systems are expanding at levels of international, regional, bilateral and multilingual cooperation.

The expansion of systems across language boundaries has led to an increase in the effort devoted to the establishment of language-independent and multilingual indexing and retrieval aids. Such aids are essential if multilingual systems are to be accessible not solely through a single, dominant language, but rather through a variety of languages, so that any indexer or searcher can work in his mother tongue, or, at least, in a language with which he is familiar. Clearly, multilingual thesauri can play a significant part in improving the control of the world's knowledge.

Guidelines for the establishment and development of multilingual thesauri were first drafted by Unesco in 1970, and revised in 1971. Since Unesco and ISO devoted the next two years to the preparation of guidelines for monolingual thesaurus construction, it was not until late 1973 that they took up the topic of multilingualism through the medium of an "International Symposium on Multilingual Thesauri", organized by ISO in Berlin (West), in close cooperation with Unesco in the framework of the UNISIST programme.

During this symposium a meeting of ISO/TC 46/WG 5 was held which elaborated a first outline, being followed by a meeting in 1974 in Helsinki where several papers were presented by experts from the Netherlands, France, United Kingdom and Federal Republic of Germany. Based on these papers, a first (new) draft of the current guidelines was prepared by G. Beling and G. Wersig (FRG) under Unesco contract in February 1975.

This paper was reviewed by the joint Unesco/ISO/TC 46/WG 5 Task Group in June/July 1976 in Paris and - due to the substantial comments - redrafted by J.A. Digger.

The International Standard ISO 2788 "Guidelines for the Establishment and Development of Monolingual Thesauri", jointly produced by Unesco and ISO in 1972/1973 and identical with Unesco's UNISIST Guidelines for the Establishment and Development of Monolingual Thesauri (SC/WS/555, Paris, September 1973), has been used as the basis for the guidelines set out in this document, on the grounds that, once a multilingual thesaurus has been established, it should be possible to divide it, if desired, into a number of independent, monolingual thesauri. All statements in ISO 2788 which relate to thesauri in general are ipso facto applicable to multilingual thesauri. Such statements will only be quoted here where this is felt to be absolutely necessary. It should be clear from the foregoing that the guidelines presented in this document are dependent on, and extend those already set out in ISO 2788. This is especially true of Section 3 in ISO 2788 (Thesaurus structure).

Any comments on this document should be sent to the Division of Scientific and Technological Documentation and Information, Unesco, 7, place de Fontenoy, 75700 Paris.

1.2 Purpose

These guidelines are intended for the ever-growing number of information services which use, or plan to use, a multilingual thesaurus. Their purpose is to describe the various tasks involved in constructing and developing multilingual thesauri, and, where possible, to recommend standard practices. To this end they draw upon the theoretical principles set out in IS 2788 and on the practical experience gained from existing multilingual systems.

There are various benefits to be had from the standardization of procedures for the establishment and development of multilingual thesauri, e.g.

- Standardization leads to compatibility between thesauri, and so may facilitate the exchange of information between different information services

- It enables those who construct or develop multilingual thesauri to learn from the experience of others, and to work in a logical manner, using standard procedures which have evolved from an international discussion of all the relevant problems.

1.3 Scope

These guidelines are purposely restricted in scope to a consideration of the problems of multilingualism as they relate to 'conventional' thesauri, i.e.

a) thesauri which are intended for use by human indexers and searchers. No mention is made of the use of automatic question analysis.

b) thesauri which are designed for coordinate (i.e. post-coordinate) indexing.

c) thesauri which avoid semantic factoring. Even in a monolingual thesaurus, semantic factoring poses special problems (see IS 2788, Section 3.2.1). In a multilingual thesaurus, therefore, it is likely to prove so difficult as to be unworkable in practice. For this reason, these guidelines will make no further mention of semantic factoring, except to suggest that the use of this technique in multilingual thesauri should be avoided if at all possible.

Throughout these guidelines emphasis will be placed on thesauri which distinguish between descriptors and non-descriptors. These may be contrasted with thesauri which treat all terms as descriptors, and which then exercise vocabulary control by converting all descriptors expressing the same concept into the same language-independent code. Thesauri of this second type will receive only passing mention here, since they do not, on the whole, pose any particularly difficult problems when used multilingually.

1.4 Limitations

Because multilingual thesauri are a comparatively recent development in the field of documentation, it is inevitable that these guidelines should, at present, display certain limitations:

a) They are mostly based on experience gained in the fields of science and technology. This means that not all of the solutions presented here may be applicable in the social sciences and humanities. Nevertheless, this draft attempts, as far as possible, to infer general principles from the specific examples it considers.

b) Any guidelines which deal with language-related issues face the problem that it is hardly possible to establish general principles which are applicable to all languages. For this reason, these guidelines concentrate on certain of the world's major languages in the fields of science and technology. Nevertheless, it is hoped that they are largely language-independent and will therefore prove applicable to languages other than those explicitly mentioned here.

2 Definitions

2.1 Definition of a multilingual thesaurus

In accordance with the general definition of a thesaurus given in IS 2788 (Section 2.1), a multilingual thesaurus may be defined either in terms of its function or its structure.

a) In terms of function, it is a terminological control device used in translating from the natural language of documents, indexers, or users into a more constrained "system language" (whose vocabulary normally forms a subset of the whole thesaurus). It differs from a monolingual thesaurus in its ability to perform this translation process in more than one language.

b) In terms of structure, it is a controlled and dynamic vocabulary of semantically and generically related terms covering a specific domain of knowledge. The concepts covered by the thesaurus are usually represented by terms of more than one natural language.+)

+) For every concept falling within its scope (or for some defined subset of these concepts), a multilingual thesaurus must provide one or more terms in each of the languages it covers. The extent to which these terms will be descriptors, as opposed to non-descriptors, will vary according to the nature of the system in use. Depending on its purpose, the system language may consist either of descriptors in each of the languages covered, or of descriptors in only one or some of the natural languages and of non-descriptors in the other.

A multilingual thesaurus is designed for both multilingual indexing and multilingual searching or for only one of these functions.
In this respect "indexing" and "searching" may include the use of a multilingual thesaurus as an auxiliary aid for finding the most appropriate description in any language.
If one of these processes is performed monolingually and the other multilingually the thesaurus may still legitimately be calles 'multilingual". However, this draft concentrates on thesauri which provide a multilingual approach to both indexing and searching. It also assumes that a property to be sought in a multilingual thesaurus is the ability to divide it, if desired, into a number of separate logically structured, monolingual thesauri.
A multilingual thesauri should not be confused with a multilingual dictionary. The two differ in certain important respects:

- the purpose of a multilingual dictionary is not to translate concepts expressed in a natural language into a restricted system language, but to show conceptual equivalences between terms in different natural languages.

- multilingual dictionaries only exert terminological control insofar as they influence usage in both technical and non-technical communication, though in no case is this done with the direct intention of aiding the retrieval of information from specific collections.

- multilingual dictionaries do not, on the whole, systematically display semantic and generic relationships between terms.

2.2 Approaches to the establishment and development of multilingual thesauri

Multilingual thesauri play an important role in international or multilingual co-operation in information work. Since co-operation in this field can take many forms, there are several different and equally valid approaches to the establishment and development of multilingual thesauri.

Before considering these approaches, a distinction needs to be drawn between the different ways in which the terms of any particular (natural) language may be used in a multilingual thesaurus. A language may be treated:

- either, as a main language, in which case some or all of the terms in this language will be descriptors, i.e. they will be terms which are directly usable in indexing, and searching,

- or, as a secondary language, in which case all terms in this language will be non-descriptors which refer indexers

and searchers to descriptors in another language. In principle, the secondary language should be incorporated into a multilingual thesaurus in such a way the thesaurus can be used by an indexer or searcher who has no knowledge of any of the main languages.

In both cases language-independent codes may be used for storage or processing purposes. The lexical and semantic structure of a multilingual thesaurus is often largely determined by the choice of main language(s).

2.2.1 Methods of use

A multilingual thesaurus may utilise main and secondary languages in various ways:

a) all languages may be designated main languages, and so be of equal status for indexing and searching purposes.

This case requires that there be a one-to-one relationship between the equivalence classes +) of all of the various languages covered by the thesaurus, i.e. if a concept is expressible in one language, provision must be made for its expression in all other languages.

b) It may use one or more main languages, plus one or more secondary languages.

2.2.2 Methods of thesaurus construction

Multilingual thesauri may be distinguished by their method of construction. The following main approaches are possible:

a) Ab initio establishment of the thesaurus. This method might be chosen by a new information system, if there were sufficient time to establish a multilingual thesaurus before the system became operational. The two methods mentioned in IS 2788 are applicable in this case:

- the a priori (Gestalt) method whereby as each new concept is considered, equivalence classes are formed in all languages. This approach is appropriate where all languages are of equal status.
- the analytical method, whereby the creation of an equivalence class in one language by indexing experiences does not necessarily lead to the immediate provision of corresponding equivalence classes in all other languages.

+) An equivalence class of terms consists of all those terms which are taken to express the same concept, i.e. a descriptor and all non-descriptors which are treated as its synonyms form a single equivalence class.

b) "Translation" of an existing thesaurus (which can be said to be in a 'source' language) into one or more other ('target') languages, the following situations can be identified:

- Translation in which feedback is permitted, leading eventually to a modification of the terms or structure in the source language. Feedback to the source thesaurus is likely to be prompted by problems encountered in the course of the translation process, and by new terminology developing in the target language. There is a tendency, with this method, for the source language to dominate the lexical and semantic structure of the resulting multilingual thesaurus.

- Translation of a thesaurus in a main language into a secondary language. In this case, it is sometimes possible to overcome difficult translation problems by adopting appropriate structures in the secondary language. Such cases have to be carefully marked.

- Descriptor-by-descriptor translation, so that every equivalence class in the source language has an exact counterpart in each of the target languages. This practice is not recommended, but may be adopted if a new language must be added to a system where it is intended that the users of this language do not, for the time being, influence the system's overall policies.

c) Merging of existing thesauri or similar vocabularies into a new multilingual thesaurus, as might be done when an international system is created on the basis of several pre-existing national systems.

3 The basic problems of multilingual thesauri

3.1 Degrees of equivalence

The basic problem in constructing a multilingual thesaurus is that of establishing equivalences between terms in different languages. Equivalences must be correctly identified if it is to be possible to convert terms from one language to another without seriously distorting their meanings. This is true irrespective of the method of thesaurus construction used, or the choice made between main and secondary languages.

When equivalences are established between terms in two languages one of these is conventionally regarded as the source language, and the other as the target language. However, in practice, there are a number of factors which tend to blur this distinction. Firstly since equivalence is a symmetrical relation, any source-target equivalence can equally well be viewed as a target-source equivalence. Secondly, the decision as to which should be the

source language and which the target language is often dependent upon the stage reached in the thesaurus construction process. Suppose, for instance, that language A is first treated as the source language and equivalences are sought in a target language B. At a later stage the roles of the two languages may well be reversed, so that B becomes the source language, and A the target language. This allows the equivalences already established between B and A to be checked, and provides an opportunity for considering any terms in B for which no equivalents in A have yet been found.

The following degrees of equivalence may be observed when a target language is examined to determine which of its terms most closely corresponds in meaning to a particular source language term:

a) exact equivalence: the target language contains a term which is identical in meaning to the source language term

b) inexact equivalence: the target language can only, at best. provide a term which is partially equivalent in meaning to the source language term

c) non-equivalence: the target language contains no term which either exactly or partially corresponds in meaning to the source language term.

These three degrees of equivalences are examined in detail in 3.1.1 - 3.1.3.

3.1.1 Exact equivalences

Descriptors which are drawn from different languages should be regarded as exactly equivalent if:

a) they have the same meaning and are also morphologically related

Example +)

English		French		German
PHYSICS	=	PHYSIQUE	=	PHYSIK

The following symbols and conventions are used in examples:

Capital letters throughout	- indicates a term which is a descriptor.
Initial letter capitalised, remainder in lower case	- indicates a non-descriptor.
=	- indicates a condition of exact equivalence between two descriptors.
= ?	- indicates that a descriptor in one language has no exact equivalent in another.
xx	- used to introduce a scope note.
UF	- abbreviation for 'Used For'.
BT	- abbreviation for 'Broader Term(s)'.

This is the most easily detected case of exact equivalence. The frequency with which this situation is met in practice will vary according to the language and subject coverage of the thesaurus.

b) they have the same meaning but are not morphologically related

Example

English		French		German
BLACKBIRDS	=	MERLE	=	AMSEL

c) they have the same denotation, but differ in connotation. This might happen, for instance, where one term is of a more technical nature than the other, e.g.

Example

English		French
BLINDNESS	=	CECITE

In cases of this kind the relationship between the descriptors is anologous to that found between quasi-synonyms in a monolingual thesaurus. Since quasi-synonyms are normally treated as equivalents (see ISO 2788, 3.4.2), it is recommended that in a multilingual thesaurus, descriptors which differ only in connotation should be regarded as exactly equivalent for indexing and retrieval purposes.

From the viewpoint of the thesaurus constructor the distinction between (a), (b) and (c) above is not of fundamental importance. In all cases, the descriptors involved have the same denotation, and can therefore be recorded as exact equivalents.

NT - abbreviation for 'Narrower Term(s)'. (Although UF, BT, and NT are English abbreviations, for convenience they are here used in French and German, as well as English, examples).

& - signifies that, for searching purposes, two descriptors are to be treated as standing in a logical product (Boolean 'and') relationship.

\+ - signifies that, for searching purposes, two descriptors are to be treated as standing in a logical sum (Boolean 'or') relationship.

3.1.2 Inexact equivalences

A classification of the types of inexact equivalence which may hold between terms in different languages will now be presented. The problems posed by each type will be considered, and various solutions suggested. To provide a context for the discussion which follows, it will be helpful to hypothesise a situation in which the aim is to construct a multilingual thesaurus containing just two languages, both of these being main languages. The two languages will be called simply 'A' and 'B'. A's descriptors will be shown as 'a_1, a_2 ... a_n', and B's as 'b_1, b_2 ... b_n'. In examples, the hierarchical level of a descriptor will be indicated by its vertical position in relation to other terms. For instance,

b_1

b_2

will indicate that b_2 is more specific than b_1. This simple natation will be used throughout this section, and in 3.1.3, 3.1.4, and 3.3.1. The use of 'A' and 'B' in preference to expressions such as 'source language' and 'target language' is deliberate. It makes it possible, for instance, to suggest generalised procedures for reconciling differences in specificity between languages which are equally applicable whether the source language is more specific than the target language, or vice versa.

It is felt that the decision to discuss inexact equivalences in terms of only two languages is justified insofar as it simplifies the guidelines but, at the same time, does not seriously detract from their usefulness. It should be clear how the techniques proposed here are extendable to circumstances in which more than two languages are used.

The solutions for each type of inexact equivalence will be categorised as follows:

1. Solutions which leave both languages unchanged
2. Solutions only involving changes to language A
3. Solutions only involving changes to language B
4. Solutions involving changes to both languages

No significance is intended in the order of the categories themselves. Within each category, the various alternative solutions will be arranged according to their order of preference. Solutions which begin by stating one or more conditions are only applicable in situations where these conditions are met.

In practice, it is common to view the solutions to equivalence problems as dividing into two broad groups:

- those which only apply if the source language may be modified on the basis of feedback from the target language

- those which are applicable whether or not feedback is allowed.

In an operational context, the four categories of solution first listed above can readily be reduced to the two broader groupings required for practical purposes:-

Source language	Categories applicable only if feedback permitted	Categories applicable whether or not feedback permitted
A	2,4	All
B	3,4	All

It may also be noted that where existing thesauri are to be merged without modification, to form a new multilingual thesaurus, only the solutions in category 1 apply.

The situations which result in ineacat equivalences between terms may be classified into four types.

a) Type 1

The nearest equivalent to a_2 in language B is b_1, but b_1 is broader in meaning than n_2, and is itself exactly equivalent to a_1.

Example

French (A)		English (B)
DIMENSIONS (a_1)	=	DIMENSIONS (b_1)
RT COTE		
COTE (a_2)	=	?
RT DIMENSIONS		

a 1. Solutions which leave both languages unchanged

Treat a_2 and b_1 as being exactly equivalent, even though they differ in specificity

Example

French		English
DIMENSIONS		
RT COTE	=	DIMENSIONS
COTE		
RT DIMENSIONS		

This solution results in a situation in which b_1 is equated with at least two descriptors in language A: a_1 and a_2. If there are other descriptors in A which, like a_2, have b_1 as their nearest equivalent, these too will be equated with b_1. The thesaurus should clearly indicate that b_1 corresponds to the combination of all terms in A which are treated as its equivalents, and that, for searching purposes, the terms within this combination are to be regarded as standing in a logical sum (i.e. Boolean 'or') relationship, e.g.

DIMENSIONS = DIMENSIONS + COTE + ...

One-many equivalences, of the kind shown above severely limit the accuracy with which the thesaurus can be used to convert:

i) index sets prepared in language B to language A
ii) search strategies prepared in language A to language B

It is clear that Solution 1 cannot be regarded as entirely satisfactory except, perhaps, in the special case where A only serves as a secondary language.

a 2. Solutions only involving changes to language A

Change the status of a_2 to that of a non-descriptor, and make a 'see' or 'use' reference from it to the broader term a_1.

Example

French		English
DIMENSIONS UF COTE	=	DIMENSIONS
COTE employed DIMENSIONS		

a 3. Solutions only involving changes to language B

a 31) Condition:

- a term which occurs as a non-descriptor in language B is equivalent in meaning to a_2.

Change this non-descriptor to a descriptor and record it as being exactly equivalent to a_2.

Example

Suppose that the French descriptor 'COTE' can be equated with the English non-descriptor 'Dimensional symbols', the English part of the thesaurus might be modified as follows:

French		English
DIMENSIONS RT COTE	=	DIMENSIONS RT DIMENSIONAL SYMBOLS
COTE RT DIMENSIONS	=	DIMENSIONAL SYMBOLS RT DIMENSIONS

a 32) Treat b_1 as a homograph by applying one of the procedures for homograph control described in 3.2.5. In this case, b_1 should be divided into two terms so that one of these retains b_1's original meaning, while the other conveys the same meaning as a_2.

Example

French		English
DIMENSIONS RT COTE	=	DIMENSIONS (SIZE) RT DIMENSIONS (Symbols)
COTE RT DIMENSIONS	=	DIMENSIONS (Symbols) xx Dimensions as represented by symbols on a plan, drawing, etc. RT DIMENSIONS (Size)

a 33) Coin a term in language B which is exactly equivalent in meaning to a_2. (For a note on coined terms see 3.1.4)

Example

French		English
DIMENSIONS RT COTE	=	DIMENSIONS RT DIMENSIONAL SYMBOLS
COTE RT DIMENSIONS	=	DIMENSIONAL SYMBOLS RT DIMENSIONS

a 34) Allow language B to adopt a_2 as a loan term

Example

French		English
DIMENSIONS RT COTE	=	DIMENSIONS RT COTE
COTE RT DIMENSIONS	=	COTE xx Dimensions as represented by symbols on a plan, drawing, etc. RT DIMENSIONS

a 35) Condition:

- It is possible to combine b_1 with one or more other descriptors in language B to form a term combination which is equivalent in meaning to a_2.

Record this term combination as being exactly equivalent to a_2. The thesaurus should indicate that, for searching purposes, the descriptors within the term combination are to be regarded as standing in a logical product (i.e. Boolean 'and') relationship.

Example

French		English
DIMENSIONS RT COTE	=	DIMENSIONS
COTE RT DIMENSIONS	=	DIMENSIONS & SYMBOLS UF Dimensional symbols Dimensional symbols use DIMENSIONS & SYMBOLS

a 4. Solutions involving changes to both languages

Conditions:

- It is possible to combine a_1 with one or more other descriptors in language A to form a term combination which is equivalent in meaning to a_2.

Record this term combination in the thesaurus. Change a_2 to a non-descriptor and make a 'see' or 'use' reference from it to the term combination. None of the descriptors in language B needs to be modified, but a record should be made of the term combination in B which corresponds to the new term combination now established in A. The thesaurus should record that, for searching purposes, the descriptors within the term combination introduced into A and B are to be regarded as standing in a logical product (i.e. Boolean 'and') relationship.

Example

French		English
DIMENSIONS	=	DIMENSIONS
DIMENSIONS & SYMBOLES UF COTE	=	DIMENSIONS & SYMBOLS UF DIMENSTIONAL SYMBOLS
Cote employez DIMENSIONS & SYMBOLES		Dimensional symbols use DIMENSIONS & SYMBOLS

b. Type 2

The nearest equivalent to b_1 in language A is a_2, but a_2 is narrower in meaning than b_1, and is itself exactly equivalent to b_2

Example

English (A)		French (B)
?	=	GROS BETAIL (b_1) NT BOEUFS CHEVAL

HORSES (a_2) = CHEVAL
BT GROS BETAIL (b_2)

b 1. Solutions which leave both languages unchanged

Treat a_2 and b_1 as being exactly equivalent, even though they differ in specificity

Example

English		French
	=	GROS BETAIL NT BOEUFS CHEVAL
HORSES	=	CHEVAL BT GROS BETAIL

This solution leads to a situation in which a_2 is equated with a combination of descriptors in B: b_1 and b_2. The thesaurus should record that, for searching purposes, the terms within this combination are to be regarded as standing in a logical sum (i.e. Boolean 'or') relationship. Mention has already been made (see Type 1, Solution 1) of the disadvantages of solutions of this type, which establish one-many equivalences between descriptors in different languages.

b 2. Solutions only involving changes to language A

b 21) Condition:

- a term which occurs as a non-descriptor in language A is equivalent in meaning to b_1

Change this non-descriptor to a descriptor, and record it as being exactly equivalent to b_1.

b 22) Coin a term in language A which is exactly equivalent in meaning to b_1. (For a note on coined terms see 3.1.4)

Example

English		French
CATTLE AND HORSES NT CATTLE HORSES	=	GROS BETAIL NT BOEUFS CHEVAL
HORSES BT CATTLE AND HORSES	=	CHEVAL BT GROS BETAIL

b 23) Allow language A to adopt b_1 as a loan term

Example

English		French
GROS BETAIL xx Cattle and HORSES NT CATTLE HORSES	=	GROS BETAIL NT BOEUFS CHEVAL
HORSES BT GROS BETAIL	=	CHEVAL BT GROS BETAIL

b 3. Solutions only involving changes to language B

Change b_1 to a non-descriptor, and make 'see', 'use' or blanket references from it to: (i) b_2, and (ii) all other descriptors in B which are immediately subordinate to b_1.

Example

English		French
		Gros bêtail Employez un terme plus précis, par example: BOEUFS CHEVAL
CATTLE	=	BOEUFS UF Gros bêtail
HORSES	=	CHEVAL UF Gros bêtail

c. Type 3

The nearest equivalent to a_2 in language B is b_1. The nearest equivalent to b_1 in language A is a_2. a_2 is narrower in meaning than b_1

Example

English (A)		French (B)
?	=	MOUTON (b_1)
SHEEP (a_2)	"	?

c 1. Solutions which leave both languages unchanged

Treat a_2 and b_1 as being exactly equivalent, even though they differ in specificity

Example

English		French
		MOUTON
	=	
SHEEP		

Repeated use of this solution may lead to a situation in which b_1 is equated with a combination of specific terms in language A. It should be recorded that, for searching purposes, the terms within this combination stand in a logical sum (i.e. Boolean 'or') relationship. Mention has already been made (see a_1) of the disadvantages of solutions of this type which establish one-many equivalences between descriptors in different languages.

c 2. Solutions only involving changes to language A

c 21) Conditions:

- a term which occurs as a non-descriptor in language A is equivalent in meaning to b_1

Change this non-descriptor to a descriptor. Change a_2 to a non-descriptor and make a 'see' or 'use' reference from it to the new descriptor.

c 22) Coin a term in language A which is exactly equivalent in meaning to b_1, taking account of the principles for coining terms given in 3.1.4. Change a_1 to a non-descriptor and make a 'see' or 'use' reference from it to the new descriptor.

Example

English		French
Mutton Use SHEEP/MUTTON		
Sheep Use SHEEP/MUTTON		
SHEEP/MUTTON UF Mutton Sheep	=	MOUTON

c 23) Allow language A to adopt b_1 as a loan term. Change a_2 to a non-descriptor and make a 'see' or 'use' reference from it to the loan term

Example

English		French
MOUTON xx Sheep and mutton UF Mutton Sheep	=	MOUTON

Mutton
Use MOUTON

Sheep
Use MOUTON

c 24) Condition:

- a_2 is a compound term which is more precombined than b_1, e.g.

German		English
?	=	TEACHER TRAINING (b_1)
LEHRERBILDUNGSGESETZ (a_2)	=	?

In some cases it will be possible to achieve a condition of exact equivalence between the descriptors by reducing the precombination level of a_2.

Example

German		English
GESETZ	=	LAW
LEHRERBILDUNG	=	TEACHER TRAINING

c 3. Solutions only involving changes to language B

c 31) Condition:

- a term which occurs as one of b_1's non-descriptors is equivalent in meaning to a_2.

Change this non-descriptor (designated b_2) to a descriptor and record it as being exactly equivalent to a_2. Identity any other of b_1's non-descriptors which, like b_2, are found to be more specific than b_1, and change these to descriptors. Change b_1 to a non-descriptor, and make 'see', 'use' or blanket references from it to: (i) b_2, and (ii) all other descriptors in B which are immediately subordinate to b_1.

Example

English		French
		Mouton Employez un terme plus précis, par exemple: MOUTON COMME BETAIL VIANDE DE MOUTON
MUTTON	=	VIANDE DE MOUTON UF Mouton

SHEEP	=	MOUTON COMME BETAIL UF Mouton

c 34) Allows language B to adopt a_2 as a loan term. Identify any of b_1's non-descriptors which are more specific than b_1, and change these to descriptors. Change b_1 to a non-descriptor and make 'see', 'use' or blanket references from it to: (i) the newly adopted loan term, and (ii) all other descriptors in B which are immediately subordinate to b_1

Example

English		French
		Mouton Employez un terme plus précis, par exemple: SHEEP VIANDE DE MOUTON
MUTTON	=	VIANDE DE MOUTON UF Mouton
SHEEP	=	SHEEP xx Mouton comme bétail UF Mouton

c 35) Condition:

- It is possible to combine b_1 with one or more other descriptors in B to form a term combination which is equivalent in meaning to a_2

Record this term combination as an exact equivalent for a_2. The thesaurus should indicate that b_1 is only to be used when it occurs as a component in the newly established term combination , or in other predefined combinations of a similar kind. It should be noted that, for searching purposes, the descriptors within these term combinations are to be regarded as standing in a logical product (i.e. Boolean 'and') relationship.

Example

English		French
		Mouton N'employez pas, mais utilisez plutôt une des combinaisons de termes énumérées ci-dessous
SHEEP	=	MOUTON & ANIMAL UF Mouton
MUTTON	=	MOUTON & VIANDE UF Mouton

c 4. Solutions involving changes to both languages

The solution so far suggested for reconciling the differences in specificity between a_2 and b_1 fall into two groups: (i) those which replace a_2 by a broader descriptor, (ii) those which replace b_1 by a more specific descriptor. The result, in each case, is a single pair of exactly equivalent terms, i.e. either:

$$a_1 = b_1$$

if a solution is selected from the first group, or

$$a_2 = b_2$$

if one is chosen from the second.

As an alternative, it is possible to retain a_2 and b_1, and to add both a new broad term to language A (as an equivalent for b_1) and a new specific term to language B (as an equivalent for a_2). The outcome will be two pairs of exactly equivalent terms:

$$a_1 = b_1$$
$$a_2 = b_2$$

The strategy is best undertaken as a two-stage process in which the 'output' from the first stage becomes the 'input' for the second. The two stages are as follows:

- Apply one of the solutions listed as b 21 - b 23 under Type 2 inexact equivalences. Each of these results in the retention of a_2, and the addition of a new broad term (a_1) to language A.

- Apply one of the solutions listed as a 31 - a 35 under Type 1 inexact equivalences. Each of these results in the retention of b_1, and the addition of a new specific term (b_2) to language B.

In practice, the order in which the stages are carried out is of no consequence, since both of the possible orders produce the same result.

Example

Inexact equivalence of Type 3

English		French
?	=	MOUTON
SHEEP	=	?

- Stage (1) - applying Solution 2(b) for Type 2 inexact equivalences

English		French
SHEEP/MUTTON NT SHEEP	=	MOUTON
SHEEP BT SHEEP/MUTTON	=	?

- Stage (2) - applying Solution 3(b) for Type 1 inexact equivalences

English		French
SHEEP/MUTTON NT SHEEP	=	MOUTON (Animal et viande) NT MOUTON (Animal)
SHEEP BT SHEEP/MUTTON	=	MOUTON (Animal) BT MOUTON (Animal et viande)

d. Type 4

The nearest equivalent to a_i in language B is b_j. The nearest equivalent to b_j in language A is a_i. a_i and b_j overlap in meaning.

d 1. Solutions which leave both languages unchanged

Treat a_i and b_j as being exactly equivalent, even though they differ somewhat in meaning

d 2. Solutions involving changes to both languages

d 21) Treat both a_i and b_j as homographs, and apply the procedures for homograph control (see 3.2.5) so that each is divided into two or more separate terms. These terms should then be examined to establish the extent of their equivalence. Exact equivalences are immediately recorded in the thesaurus. Inexact equivalences of types 1-3 are resolved by adopting one of the solutions described above. Non-equivalences are handled according to the procedures suggested in 3.1.3.

d 22) Coin a new term for language A which is exactly equivalent in meaning to b_j. Similarly, coin a new term for language B which is exactly equivalent to a_i. (For a note on coined terms see 3.1.4).

d 23) Allow language A to adopt b_j as a loan term. Similarly, allow language B to adopt a_i as a loan term.

3.1.3 Non-equivalences

A situation of non-equivalence occurs when one language (A) contains a descriptor (a_1) for which another language (B) has no direct or inexact equivalent. Two solutions to this problem can be suggested:

- Coin an artificial term which can be added to language B as an exact equivalent for a_1. (For a note on coined terms see 3.1.4).

Example

English	French
STEAM CRACKING	VAPOCRAQUAGE xx Craquage à la vapeur d'eau

- Allow language B to adopt a_1 as a loan term. This may be the only possible course of action in the case of terms which are commonly thought of as being 'untranslatable', e.g. KISMET (Arabic), GEMÜTLICHKEIT (German), HIPPY (English).

In general, no clear preference can be given between the solutions outlined above. When deciding on the most appropriate method of adding a new term to a language, three main factors should be taken into account:

- the normal practice of the users of that language with regard to the adoption of artificial terms and loan words
- the language ability of the users of the thesaurus
- the extent to which the terminology of the subject field covered by the thesaurus is dominated by any particular language or group of languages

3.1.4 Establishing equivalences

The analysis of equivalence problems presented above provides a framework for establishing equivalences between the descriptors in two languages according to a logical sequence of operations. This sequence can best be demonstrated by again referring to two hypothetical languages, A and B. It will be assumed that the process of thesaurus construction has reached a stage in which A is to be treated as the source language. The steps involved in seeking an equivalent in B for a given term in A (say a_2) are as follows:

a) Examine B for a term which is identical in meaning to a_2.

- If such a term can be found record it as being exactly equivalent to a_2

- If not, proceed to 2.

b) Examine B to determine which, if any, of its terms are close in meaning to a_2.

- If only one such term can be found proceed immediately to 3.
- If two or more such terms can be found, identify one of these as being the nearest equivalent to a_2, and then proceed to 3.
- If no such term can be found, a_2 is to be regarded as having no exact or inexact equivalent in B. One of the solutions for cases of non-equivalence should, therefore, be applied (see 3.1.3).

c) Identify the type of inexact equivalence which hold between a_2 and its nearest equivalent in B (see 3.1.2). Proceed to 4.

d) Select from among the solutions for the appropriate type of inexact equivalence one which is applicable to the case under consideration.

Step 2 and 4 call for special comment:

- Step 2: If a language contains two or more terms which are close in meaning to a given term in another language, one of these has to be chosen as the nearest equivalent. Until this is done it is impossible to identify the type of equivalence problem for which a solution is sought. Consider, for instance, the following parallel hierarchies in French and English:

French		English
BETAIL	=	LIVESTOCK
GROS BETAIL	=	?
BOEUFS	=	CATTLE

If LIVESTOCK is taken as the nearest equivalent to GROS BETAIL, the equivalence problem can be identified as that described in 3.1.2 as Type 1 (where French corresponds to the language called 'A' in the explanation of Type 1, and English corresponds to 'B'). If, on the other hand, CATTLE is treated as the nearest equivalent, the problem changes to one of Type 2 (in this case, French = B, English = A).

In practice, if several terms in one language (say B) are close in meaning to a given term in another (a_2), the term chosen as a_2's nearest equivalent will be either:

a) the term which is closest in meaning to a_2, or, if no one term can be found which is clearly closer in meaning than the others

b) the term whose inexact equivalence with a_2 can most easily be reconciled

- Step 4: The choice of solution is dependent on a number of factors:

  a) the extent to which A and B may be modified, and, in particular, the extent to which A (the source language on this occasion) may be changed in the light of feedback from B (the target language). This determines which categories of solution are applicable.

  b) any conditions peculiar to the case under consideration, which limit the range of applicable solutions within any given category.

  c) the order of preference indicated among the applicable solutions.

Certain of the solutions proposed in 3.1.2 are based on the notion that a non-descriptor may be changed to a descriptor to allow it to serve as an exact equivalent for a descriptor in another language. Leaving aside this special case, non-descriptors do not normally enter into the process of establishing equivalences between terms in different languages, except insofar as they may, by their presence in a thesaurus, help to define the scope of ambiguous descriptors. It is recommended that when the construction of a multilingual thesaurus is undertaken, effort should be concentrated on establishing equivalences between descriptors. The further work required to find correspondences between all non-descriptors is hardly likely to be justified in terms of the additional benefits to be gained.

On occasions, the only practical solution to a case of inexact or non-equivalence between two languages (A, B) is to coin a term which can be added to one of the languages (say B) as an exact euqivalent for a term (a_1) in the other. If the concept conveyed by a_1 is directly expressible in B, then a term can often be created simply by translating each of the elements of a_1 into B, e.g.

German term (a_1)	Coined English term
SCHÄFERHUNDERZIEHUNG	SHEEP DOG TRAINING

Coined terms constructed in this way, i.e. by a process of literal translation, often appear perfectly 'natural', being simply compound expressions which, though originally omitted from B's indexing and retrieval vocabulary, might equally well have been included. In other cases, there may be no way of avoiding a coined term which is a genuine neologism - as may happen, for instance, if the literal translation of a_1 produces an artificial term. Such a term can be accepted as a satisfactory coined term, provided it is intelligible to the users of language B.

Example

English term (a_1)	Coined French term
STEAM CRACKING	VAPOCRAQUAGE
SERVICE STATION	STATION SERVICE

As the example demonstrates, it may sometimes be desirable to modify the literal translation of a_1 so as to simplify the form of the coined term, i.e. CRAQUAGE A LA VAPEUR D'EAU is reduced to the simpler VAPOCRAQUAGE!

If the difference in strcture between language A and B is such that literal translation between the two is impossible, no attempt can be made to coin a term in B which preserves the syntax or morphology of a_1. In cases of this kind, consideration should be given to the relationship between a_1 and the terms in A which might be calles its 'semantic neighbours'. It may be possible to coin a term in B whose equivalence in meaning with a_1 is shown by the fact that the two terms relate to their semantic neighbours in a similar way. Consider, for instance, the terms LOGICIEL and MATERIEL which are the French equivalents for the English terms SOFTWARE and HARDWARE. The decision to coin these particular terms was based on the fact that they contrast in meaning in much the same way as their English counterparts.

A term may be coined by combining two or more existing descriptors, if, by this means, it is possible to provide a new term at an appropriate level of specificity, e.g. SHEEP/MUTTON (see 3,1.2, c. 4).

If a_1 has no exact or inexact equivalent in B, this may indicate that it expresses a concept which is peculiar to the culture of the users of language A (or the language group to which A belongs). No attempt should then be made to equate a_1 with a term in B which, though different from a_1 in denotation, is regarded as having analogous connotations. For instance, it would obviously be wrong to treat the Icelandic word for 'fish' as being equivalent to the English word 'bread' when this is used as the symbol of food in general. Clearly, this particular example is hypothetical, insofar as the word 'bread' would only be used in the sense indicated in natural language, and not in a documentary language. However, the principle involved is equally valid for documentary languages such as multilingual thesauri.

In general, the thesaurus compiler should regard the use of a coined term to reconcile inexact equivalences and non-equivalences between languages as a solution of last resort, and not as a panacea for all problems of this type. Wherever possible, a coined term should be viewed as a temporary descriptor, i.e. it should be replaced as soon as it is superseded by a synonymous term which has greater literary warrant.

A coined term which is only added to a language to facilitate the establishment of a multilingual thesaurus, and which is more

specific than the users of that language require, may be handled in a special way. It may be retained in its fully specific form for the purposes of providing services to a multilingual user population, but automatically converted to a broader descriptor where only monolingual access is to be offered. In these circumstances, since the coined term is never required to function as a search term (being either converted to a broader descriptor in the same language, or to an equivalent descriptor in another language) it may take the form of an arbitrary 'concept number'.

Example

French		English
DIMENSIONS RT COTE	=	DIMENSIONS RT 1234 (Dimensional symbols)
COTE RT DIMENSIONS	=	1234 UF Dimensional symbols Dimensional symbols use 1234

3.1.5 The adjustment of term relationship

As 3.1.2 and 3.1.3 demonstrate, various modifications may be made to a language in the course of applying the solutions for reconciling inexact equivalences and non-equivalences between terms:

- new descriptors may be added
- non-descriptors may be changed to descriptors
- descriptors may be changed to non-descriptors

Each of these changes affects the structure of the vocabulary and should be reflected in an adjustment to the relationship shown between terms, whether these be displayed by systematic listings, references (e.g. BT, NT, etc.) arrowgraphs, or by some other means.

If a new descriptor is added to a thesaurus, or if any existing non-descriptor is made into a descriptor, the new term should be related to:

a) its synonyms and quasi-synonyms. Any of these which are not already present in the thesaurus should be added as non-descriptors

b) any other descriptors which are related to it in a hierarchical or associative manner

Example:

Suppose the coined term CATTLE AND HORSES is added to the English part of a multilingual thesaurus as an equivalent for the French GROS BETAIL. The effect of this addition on

the display of term relationship might be as follows:

Original relationship	Relationship after addition of CATTLE AND HORSES
	CATTLE AND HORSES BT LIVESTOCK NT HORSES - etc -
HORSES BT LIVESTOCK	HORSES BT CATTLE AND HORSES
LIVESTOCK NT HORSES - etc -	LIVESTOCK NT CATTLE AND HORSES - etc -

This example demonstrates that the relational connections established for new descriptors (e.g. BT LIVESTOCK and NT HORSES for CATTLE AND HORSES) may supersede some of those already recorded for other terms (e.g. BT LIVESTOCK for HORSES, and NT HORSES for LIVESTOCK). Superseded connections of this kind should always be deleted from the thesaurus.

If a descriptor is changed to a non-descriptor its relational connections should be carefully reviewed. All connections which are retained must be changed so that they indicate equivalence relations (e.g. 'Use/UF'), this being the only type of relation a non-descriptor may hold with another term. Any relational connections which are found to be no longer applicable to the term in its capacity as a non-descriptor should either be discarded, or, preferably, transferred to an appropriate descriptor.

Example

Original relationships	Relationships after GROS BETAIL changed to a non-descriptor
BETAIL NT GROS BETAIL	BETAIL UF Gros bêtail NT CHEVAL
CHEVAL BT GROS BETAIL	CHEVAL BT BETAIL
GROS BETAIL BT BETAIL NT CHEVAL	Gros bêtail employez BETAIL

3.2 The elements of a multilingual thesaurus

3.2.1 Descriptors

The monolingual sub-thesauri which collectively make up a multilingual thesaurus should be alike in structure. It would be inadmissable, for instance, to allow one sub-thesaurus to distinguish between descriptors and non-descriptors, but have the others treat all terms as being of equal status. Every multilingual thesaurus should be recognisably of one of the two thesaural types identified in ISO 2788 (Section 2.1).

It is generally recommended that each descriptor in any sub-thesaurus which uses preferred terms should have at least one equivalent term in each of the other sub-thesauri. This rule may sometimes be waived in the case of secondary languages. In choosing a descriptor from among a set of terms which form an equivalence class, every effort should be made to select the term which has the greatest degree of international acceptance, i.e. preferably, a term of Latin or Greek origin. This is particularly important in the case of general terms and facilitates international communication and understanding.

Example

English	German
TELEPHONE	TELEFON NF Fernsprecher

Descriptors and non-descriptors in multilingual thesauri form equivalence classes for each language but for the scope of the thesaurus, too. This may lead to the use of numbers or similar artificial language symbols to represent the class of equivalent descriptors, i.e. the concept. Such an approach is most likely to be adopted in computer-aided storage and retrieval systems.

3.2.2 Proper names

The form of a proper name may vary from language to language:

a) A name used in the source language may occur in the target language in both its original and translated forms but the original being the more common usage.

Example

English	German
NEW YORK	NEW YORK UF Neuyork

In such cases the untranslated form of name should be adopted as a descriptor in the target language, the translated may be added as a non-descriptor.

b) A name in the source may commonly occur in the target language in translated forms. Here, the form which is closest to the original (untranslated) name should be adopted as a descriptor in the target language.

Example

English	German
BIELORUSSIAN REPUBLIK	BJELORUSSISCHE REPUBLIK UF Weißrussische SSR

c) Because of differences in the various standards for transliteration and transcription, a name can sometimes occur with alternative spellings. In general, the form of name chosen to serve as a descriptor should conform to the spelling recommended by ISO standards. References should be made to the preferred form of name from all common variants - these being treated as non-descriptors. Where no ISO standards exist, or where they exist but conflict with national standards, the latter should be used to determine the preferred form of name and the source be indicated.

In the absence of any applicable standards, whether national or international, the form of name adopted as a descriptor should be that which commands the greatest degree of international acceptance. Common national variants of the preferred form of a name should be treated as non-descriptors.

In terms of the rule set out in Section 3.2.2.b. above, geo-political regions are often incorrectly named in thesauri. Such errors might be thought trivial in a monolingual thesaurus, but are serious defects in a multilingual thesaurus which is intended for international use. Hence, when dealing with the names of geo-political regions, every care should be taken to select the correct form of name as a descriptor. References should be made to this form from all incorrect forms of name which are in common usage.

3.2.3 Abbreviations and acronyms

In general, abbreviations and acronyms should be avoided as is indicated in IS 2788 (Section 3.2.7). If an abbreviation or acronym can take different forms in different languages, the form used in the source language should be entered in the target language thesaurus as a non-descriptor, and accompanied by an appropriate scope note.

3.2.4 Noun form and numbers

The preference for using descriptors in the noun form (as recommended in ISO 2788, Section 3.2.4) sometimes raises problems.

Example: The English term "painting" refers both to the process of painting (German: 'das Malen') and also to the product of this process (German: 'das Bild')

A polysemic descriptor of this kind can be treated in a thesaurus in the same way as homographs (s. 3.2.5).

The example above shows that the decision as to the most appropriate form of descriptors for a particular concept may be language-dependent (see ISO 2788, Section 3.2.5). Hence, two descriptors in different languages may have different grammatical forms, and yet be treated as semantically equivalent.

Example

English	German
SAWS	SAERGE

3.2.5 Homographs and polysemes

From a strictly linguistic viewpoint homographs and polysemes are quite distinct, but to the indexer and searcher they pose exactly the same problem - that of a term which can be ascribed with more than one meaning. For this reason these guidelines will adopt 'homograph' as a convenient expression to cover instance of both homography and polysemy.

Unless homographs are adequately controlled they can be a source of confusion in a thesaurus. This is particularly true of multilingual thesauri, where homographs can occur in two forms.

a) Intra-language homographs. Where the thesaurus compiler encounters a term of this kind he may inadvertently treat it as being equivalent to two or more terms in another language. Appropriate steps need to be taken to detect such cases and to split the homograph into two or more separate terms. The following methods may be used to distinguish between the different meanings of an intra-language homograph.

a 1 Replace the homograph by two or more separate terms, each of which conveys only one of the homograph's meanings

English example

FILMS, meaning either 'thin films' or 'photographic films',
can be replaced by: THIN FILMS
PHOTOGRAPHIC FILMS

a 2 Attach verbal qualifiers to the homograph to make each of its various meanings explicit.

French example

MOUTON, which refers either to the animal or to its meat, can be expanded to: MOUTON (animal)
MOUTON (viande)

a 3 Use the homograph in both its singular and plural forms, if each of these conveys a different meaning

English example

WOOD, which refers either to a type of material or to a tree-covered area of land, can be divided into:

WOOD
xx Fibrous substance obtained from trees and used as raw material of fuel

WOODS
xx Tree-covered areas of land

a 4 Retain the homograph, but restrict its scope to a single meaning. Use synonyms to convey the homograph's other meaning.

German example

SEE, meaning either der See (lake) or die See (sea), may be divided into:

SEE
xx Der See

See, die benutze MEER
MEER
UF See, die

a 5 Attach numbers or other arbitrary symbols to the homograph so as to split it into two or more separate terms of differing scope

English example

GOLF CLUBS, which refers either to a type of sports club, or to an implement used in playing golf, can be divided into:

GOLF CLUBS 1
xx Sports clubs for golfers

GOLF CLUBS 2
xx Implements used in playing golf

Not all of the methods described above are generally applicable - for instance, (i) only applies where a separate term can be found for each of the homograph's meanings. The methods are arranged in order of preference. It is recommended that when a homograph is encountered it should be handled by selecting whichever of the applicable methods of homograph control is shown as having the highest preference rating. Note that methods (iii), (iv) and (v) require the use of scope notes.

b) Inter-language homographs, i.e. terms (whether common nouns, proper names, or abbreviations) which occur in two or more languages but which have different meanings in each. Inter-language homographs should always be marked with language indicators. If this is insufficient to remove the possibility of confusion, scope notes should also be added.

Example

English		German
ANGER(E)	=	ZORN
MEADOWS	=	ANGER (D)

where, '(E)' and '(D)' indicate 'English' and 'Deutsch', respectively.

Any term which occurs as both an intra- and inter-language homograph should first be subjected to one of the procedures for the control of intra-language homographs. If, as a result, the form of the term is changed, so that it no longer stands as an inter-language homograph, no further action is required. Otherwise, it should be marked with appropriate language indicators.

3.2.6 Definition and scope notes

It should perhaps be stated that there are certain situations in which the inclusion of a definition is almost mandatory e.g.

- when a term is used in both its singular and plural forms
- when the term is qualified by an arbitrary symbol (e.g. STANDARDS[1]).
- when an artificial term has been coined in the target language in order to provide an equivalent for a term in the source language, e.g.

'Educational law' as
an equivalent for 'Bildungsgesetz'

- when a loan term has been added to a language.

Sometines, although two descriptors in different languages are equivalent, one requires a definition or scope note, while the other does not. This situation may arise because:

a) Only the source language descriptor is ambiguous and requires definition. In this case, a definition or scope note need not be attached to the target language descriptor. This omission should be noted at least in the process of constructing the thesaurus.

Example

English	German
WATER COOLING	WASSERKÜHLUNG
Cooling with water	%
	% = scope note omitted

b) Only the target language descriptor requires definition. Steps should be taken to ensure that an appropriate definition is included in the thesaurus entry for this term, particularly if both source and target languages are to be used as main languages.

Example

English	German
CHEMURGY	CHEMURGIE § Befaßt sich mit der Verwendung von Agrar- und Forstprodukten in der chemischen Industrie

c) The definition attached to the source language term is considered inappropriate for the target language term, either because it is difficult to understand, or because it is non-standard. In this case, the definition should be altered. This alteration should be noted in the thesaurus, together with the source of the new definition supplied.

3.3 Relationships

3.3.1 Hierarchical relations

By 'hierarchical relations' is here meant all relations - generic and partitive - which are used to establish levels of subordination and superordination among descriptors. If a hierarchy in one language (A) is compared with the corresponding hierarchy in another (B), the two will often display one or more differences in structure. In the simplest case, a difference of this kind will not be due to a lack of equivalence between terms, but will simply reflect the fact that one language (say A) shows a particular hierarchical relation between terms, which the other (B) does not. If the extra hierarchical relation present in A is specific to the culture of that language's users, it cannot validly be added to B; the two languages must be allowed to differ in structure. Conversely, if this hierarchical relation is not language-dependent it may be added to B, provided this does not contravene any rules limiting the use of feedback from one language to the other.

In the more complex case, the presence of an additional hierarchical relation in A indicates that, at one point in the hierarchy, A contains a descriptor which has no exact equivalent in B. Hierarchical differences between languages which are due to inexact equivalences and non-equivalences between descriptors can be eliminated, or at least normalised, by bringing the descriptors in question into exact equivalence. Four common types of hierarchical difference between languages are considered below. Each is traced to a particular type of equivalence problem (as defined in 3.1.2 - 3.1.3), so that a solution appropriate to this problem may be used to reconcile the difference between the hierarchies.

a) A descriptor in A which occurs at an intermediate position in a hierarchy has no exact equivalent in B. This situation might be represented as follows:

$a_1 = b_1$

$a_2 = ?$

$a_3 = b_3$

The first requirement is to find the nearest equivalent to a_2 in B. If b_1 is taken as the nearest equivalent, the equivalence problem is of Type 1 (see 3.1.2). One of the solutions for Type 1 should be used to reconcile the difference in specificity between a_2 and b_1, and so bring the two hierarchies

into equivalence. If, on the other hand, b_3 is regarded as a_2's nearest equivalent, the equivalence problem then corresponds to Type 2 in 3.1.2, and the choice of solution should be made accordingly.

b) A descriptor in A (a_i) and a descriptor in B (b_j) overlap in meaning, but have equivalent sets of narrower terms.

One of the solutions for handling terms of overlapping meaning (see 3.1.2, Type 4) should be applied. If the solution adopted leads to changes in the descriptors of A and B, the term realtionships within these languages should be checked, and where necessary adjusted, so as to ensure that the descriptors which were previously recorded as narrower terms for a_i and b_j are now subordinated to the more appropriate broader terms.

c) A descriptor in A (a_1) and a descriptor in B (b_1) are exactly equivalent, but have different sets of narrower terms.

Each of a_1's narrover terms which has no exact equivalent in B should be handled by applying one of the solutions for equivalence problems of Type 1 (see 3.1.2). b_1's narrower terms which lack exact equivalents in A should be treated in the same way.

d) A descriptor in A (a_i) and a descriptor in B (b_j) overlap in meaning, and have different sets of narrower terms. This is a combination of situations (b) and (c) above.

The terms of overlapping meaning should be considered first. These should be reconciled by applying one of the solutions for equivalence problems of Type 4 (see 3.1.2). Attention should then be turned to the narrower terms in A and B which lack equivalents in the other language. Each of these should be handled by applying one of the solutions for equivalence problems of Type 1 (see 3.1.2).

3.3.2 Associative relations

The display of associative relations (RTs) in a thesaurus helps to indicate how terms are defined. It has already been noted that a multilingual thesaurus may sometimes treat descriptors as though they were equivalent even when strictly they differ somewhat in meaning. If follows that the decision as to what associative relations are appropriate in a particular language should, in the first instance, be made in the light of that language's special requirements. At a second stage - provided feedback of this type is permitted - the language may have added to it any further associative relations suggested by other languages. RTs added to a language in this way should be marked to indicate their 'language of origin'.

4 Presentation and arrangement

4.1 The introduction to the thesaurus

In overall content, the introduction to a multilingual thesaurus should meet the general requirements set out in IS 2788, Section 4.1. In addition, it should make specific mention of the following points:

a) the nature of the multilingual thesaurus (see Section 2.2)

b) the method of reconciling the discrepancies between hierarchies of terms in different languages, together with a list of any special symbols (such as language indicators) used for this purpose

c) the method of establishing equivalences between equivalence classes in different languages.

It is essential that the introduction be written in all of the languages covered by the thesaurus, including all secondary languages.

4.2 The main part of the thesaurus

The organization of the main part of the thesaurus (whether alphabetic or systematic in its arrangement) should generally conform to the principles set out in IS 2788. Within the main part the entries should show all the equivalent descriptors in the other languages covered by the thesaurus.

If a thesaurus contains two or more main languages, each must be given a main part. Where there is more than one main part these may be arranged as follows:

- separately, i.e. a separate main part is provided for each language

- in parallel, i.e. the main parts for the various languages are presented systematically in parallel columns on the page. Where this method can be used it is preferable, since it results in an arrangement of terms which is much more easily scanned by the thesaurus user than that produced by the provision of separate parts.

4.3 Auxiliary parts of the thesaurus

4.3.1 Alphabetical indexes

The general requirements for alphabetical indexes, as set out in IS 2788, are equally applicable for multilingual thesauri. Each language requires an alphabetical index if the main part is arranged systematically or if this language has no main part. Alphabetical indexes to the various languages covered by a multilingual thesaurus may be presented in the following ways:

a) separately, i.e. in the index for any particular language, the entry for each term simply refers to that term's location within the main part of the thesaurus. This method has the virtue of being cheap to implement, but is only recommended if the main part is itself arranged in parallel or the secondary language terms are treated as synonyms to the main language terms

b) in parallel, i.e. in the index for any particular language, the entry for each term not only refers to that term's location within the main part, but also records equivalent terms in one or more other languages. Each language may be provided with:

- either, a comprehensive parallel index, in which each entry lists the equivalent terms in all of the other languages covered by the thesaurus
- or, a dual parallel index, in which each entry lists the equivalent term(s) in only other language. In this case, indexes are required for all possible pairs of languages

Example: If a systematic multilingual thesaurus covers English (E), French (F), and German (D), the following dual parallel indexes have to be constructed: E-F, F-E, E-D, D-E, F-D, D-F.

4.3.2 Systematic listings and graphic displays

The provision of a systematic listing may raise severe problems, or even prove impossible, if the terms in the various languages are to be listed in parallel. The compilation of a systematic listing generally calls for much greater care and effort in a multilingual thesaurus, than it does in a monolingual one.

Parallel systematic listings are clearly more useful than separate ones, particularly if hierarchies are represented by graphic displays. These can incorporate all the relevant information about the scope and use of each of the descriptors in the various languages.

4.4 The relationship between the main and auxiliary parts of the thesaurus

The general guidelines on the arrangement of thesauri, as given in ISO 2788, are here elaborated to take account of the special needs and problems of multilingual thesauri.

If a thesaurus contains just one main language, and one or more secondary languages, only the main language need be provided with a main part, all secondary languages being at least with an alphabetical index.

5 The process of constructing a multilingual thesaurus

5.1 Preliminary procedures

5.1.1 Notification of appropriate clearinghouses

Any information system or agency which intends to construct a multilingual thesaurus should notify the appropriate international clearinghouse(s) of its intention. This information should be channelled to the clearinghouse(s) through whatever national organization(s) are concerned with thesauri and their construction. Once a thesaurus has been produced, a copy of both the first edition and all subsequent editions should be sent to the clearinghouses previously notified. These clearinghouses should be listed in the introduction to the thesaurus.

5.1.2 Avoiding duplication of effort

Before a new multilingual thesaurus is constructed for a particular subject field, every effort should be made to discover whether a monolingual thesaurus already exists in this field or in any cognate subject area. This information can be obtained from the relevant thesaurus clearinghouses +). An approach should also be made to INFOTERM ++) to discover what glossaries are available for the subject field in question.

If a multilingual thesaurus is to be constructed ab initio, in a subject area in which one or more such thesauri already exist, all existing thesauri and sources of terminology (in particular ISO vocabularies) should be taken into account when preparing the new thesaurus. The introduction to the new thesaurus should explain why it was constructed and how it differs from existing documentation languages in the field.

5.2 Establishing equivalence classes

This involves three stages:

a) defining the objectives of the thesaurus and its intended method of use, and deciding which languages are to be treated as main languages and which as secondary.

+) e.g. Institute for Scientific, Technical and Economic Information, The Clearinghouse (IINTE). Al Niepodleglosci 186, Warszawa, Poland

++) Internationales Informationszentrum für Terminologie (INFOTERM), affiliated to Austrian Standards Institute, Leopoldsgasse 4, A 1041 Vienna 2, Austria

b) defining the method of construction

c) collecting and evaluating terms, and establishing equivalence classes. A high degree of international and inter-lingual co-operation is essential for the successful accomplishment of this stage.

The activities which take place at these three stages vary according to the method of thesaurus construction.

5.2.1 Ab initio method of construction

If a multilingual thesaurus is to be constructed ab initio three situations can be distinguished:

- One of the languages to be covered by the thesaurus is selected as the main language, and the remainder are made secondary languages. In this case, the main language can be used as the skeleton of the multilingual thesaurus. The method of construction will resemble that used when a multilingual thesaurus is produced by the translation of a monolingual thesaurus (see Section 5.2.2), with the difference that the source thesaurus for the main language must, in this instance, be created ab initio before translation can begin. This thesaurus should be established according to the principles defined in IS 2788.

- The terms for each language are collected separately, in different countries, or by different agencies. In this case, the method of construction will resemble that used when existing monolingual thesauri in different languages are merged (see Section 5.2.3).

- The terms for each language are collected simultaneously by a single organization.

The remainder of this section will concentrate exclusively on the last of the three methods of construction outlined above, since this is the only method which can truly be said to be based on an ab initio approach.

Firstly, the overall structure of the multilingual thesaurus must be defined, and a choice made between the (a priori) Gestalt and analytical approaches. In either cases, the procedure for collecting terms is essentially as for monolingual thesauri (see Section 5.5.1). Terms may be collected from a variety of sources, but preference should be given to existing multilingual sources where these are available. The extraction of terms from a multilingual source should proceed in a manner which immediately establishes equivalences between terms in different languages; it is inadvisable to extract the terms for each language separately. Monolingual sources of terms may also prove useful and should be carefully studied.

For control purposes, a unique number (preferably incorporating a check digit) may be assigned to each of the terms gathered during the preliminary stages of thesaurus construction.

Secondly, the corpus of terms collected should be divided into subsets of such a size that they can easily be scanned and processed in a consistent manner. The terms within each subset should be grouped into equivalence classes each of which may be given a unique concept number. Initially, it is not essential that all equivalence classes contain (at least) one term from each language. If the analytical approach is used 'gaps' in equivalence classes can be filled as new terms are added. With the Gestalt approach, equivalence classes can be completed at the time the relationship between them are established.

The stages in the ab initio method of thesaurus construction can be summarised as follows: divide the vocabulary into subsets of terms, identifying the equivalence classes within each subset and record the relationships between these classes. The end product of these activities will be the first draft of a multilingual thesaurus.

5.2.2 "Translation" of an existing (source) thesaurus into a target thesaurus

It is recommended that, where possible, translation should proceed from a monolingual source thesaurus. If an additional language is to be added to an existing multilingual thesaurus it is preferable to select only one language as the source language. The simultaneous conversion of terms in several source languages into one target language increases the number and complexity of the translation problems and seldom provides better results than those obtained from a single source language. The source language should be chosen from among the dominant languages in the literature of the subject field to be covered by the thesaurus.

The translation method of constructing a multilingual thesaurus is normally based upon the Gestalt approach. However, should the analytical approach be chosen instead, the procedures to be followed will resemble those adopted when the ab initio method of thesaurus construction is applied to produce a multilingual thesaurus with one main language.

It is unlikely that all the terms in the source thesaurus can be translated by a small group of subject specialists. Hence to prepare it for translation, the thesaurus will need to be divided into groups of terms, each of which can be translated separately. Two distinct approaches are possible:

- translation of hierarchy-by-hierarchy. The source thesaurus is divided into its constituent hierarchies, each of which is supplied to a translator who trans-

lates only descriptors, and ignores all further references as related terms. Although this method of translation often results in an excellent choice of descriptors for the target language, it brings with it the disadvantage of problems caused by polyhierarchical relationships: terms which occur in more than one hierarchy will be translated more than once. Careful control procedures will be required to deal with this situation.

- descriptor-by-descriptor translation. Each translator is supplied with a group of terms, each term in the group being accompanied whereever possible by a list of the terms to which it is related. The sole purpose of this list is to aid the translator by elucidating the meaning of the term to be translated. This method ensures that each term is translated once only. However, it may lead to situations in which a source term is translated into only one term in the target language when it ought, ideally, to have been translated into two or more. Errors of this kind are likely where the translator has failed to notice that the source term in question is a member of more than one hierarchy, and that its meaning varies according to its hierarchical context.

Both approaches require that each descriptor in the source language be translated into at least one corresponding descriptor in the target language. The kinds of equivalence which may be achieved between source and target language descriptors have already been described in Section 3.1. Once the source thesaurus has been translated to a target language, the resulting terms and hierarchies should be amalgamated to provide the first draft of the multilingual thesaurus.

5.2.3 Merging of existing thesauri

The merging of two or more separate thesauri to produce a new multilingual thesaurus is a time-consuming and error-prone task involving the reconciliation of various kinds of non-equivalence between terms and hierarchies (see Section 3.1). For this reason, this method of constructing a multilingual thesaurus cannot be recommended, until further experience in this method is gained.

5.3 Control procedures

It is recommended that before any attempt is made to establish equivalences between terms in different languages, vocabulary control should first be exercised within each language, so that synonyms and homographs are controlled and each equivalence class is identified by a unique descriptor. This will greatly simplify the task of preparing the first draft of a multilingual thesaurus. Once this draft has been completed it should be examined to check the equivalences established between

the descriptors in the different languages. Strictly, these equivalences are established between equivalence classes, but since, in indexing, each equivalence class is represented by a descriptor, it is permissible to talk of equivalences between descriptors. Occasionally, an equivalence class may be represented not by one descriptor, but by a term combination - see, for instance, the French term combination for 'sheep' shown in the example for Solution 3 (e), Type 3, 3.1.2, i.e.

MOUTON & ANIMAL.

Hence, although this section will use the simple expression 'descriptor' throughout, this should be interpreted as meaning: any descriptor or combination of descriptors used to represent an equivalence class . The establishment of equivalences between non-descriptors will not be discussed, since, as previously noted (see 3.1.4), the additional work which this entails can rarely be justified in terms of the extra benefits to be expected.

To make the checking of equivalences easier, computer assistance can be used. Typically, this will involve the following procedures:

- Each of the concepts falling within the scope of the thesaurus is given a unique 'concept number'.
- Each descriptor is assigned the number of the concept is expresses
- Descriptors, plus their accompanying concept numbers, are converted to machine-readable form to produce a 'descriptor file'.
- A computer processes this file to generate a multilingual list showing all the descriptors pertaining to each concept number.

In order to consider how the problems and errors revealed by the computer-generated list should be handled, it will again be helpful to imagine a situation in which only two languages are in use. For any given concept, the list may show any one of the following types of correspondence between the descriptors in the two languages:

a) one descriptor is listed for each language, i.e. there is a one-to-one correspondence between the languages

b) one descriptor is listed for one of the languages, but none is listed for the other, i.e. a one-to-none correspondence

c) one descriptor is listed for one of the languages, but two or more are listed for the other, i.e. a one-many correspondence

d) two or more descriptors are listed for each language, i.e. a many-many correspondence

e) two or more descriptors are listed for one language, but none is listed for the other, i.e. many-to-none correspondence

In case (a), it is reasonable to assume that the descriptors listed for the two languages are exactly equivalent. This assumption should, however, be verified, before the descriptors are entered in the final version of the multilingual thesaurus as exact equivalents, since the possibility that a one-to-one correspondence between terms has been established in error cannot be entirely disregarded.
If the two descriptors are found to be only partially equivalent in meaning, the methods for reconciling this situation should be determined by reference to the classification of inexact equivalences presented in 3.1.2. If the two descriptors are found to be totally non-equivalent, the procedures for establishing equivalences (see 3.1.4) should be applied to each.

Cases of type (d) and (e) can be traced to one or both of the following factors: (i) an error in the allocation of concept numbers - such as the assignment of the same number to more than one concept; (ii) an error in vocabulary control leading to a situation in which one of the languages, or perhaps both of them in the case of (d), contain synonymous descriptors. The first step in handling (d) and (e) is to identify the errors involved, and to correct these, e.g. by re-allocating concept numbers, and by merging all synonymous descriptors and their associated equivalence classes into a single descriptor representing a single equivalence class. As a result, each occurrences of (a), (b) or (c), i.e. the elimination of errors should ensure that, for each concept, there is at least one language for which only one descriptor is listed.

Since (d) and (e) can be converted to (a), (b) or (c), and since (a) poses no particular problems, only (b) and (c) require further explanation. Step-by-step procedures for handling these cases are described below. For the purposes of this description the language which the computer-generated list shows as having just one descriptor for a given concept will be called A, and the descriptor listed for this language will be called a_1. The other language which, for the same concept, is shown as having either more than one descriptor (case (c)), or else no descriptor at all (case (b)), will be called B. The following procedures are applicable when considering the descriptors listed for a given concept:

a) <u>One-to-none correspondences - Case (b)</u>

a 1. Check for errors in the allocation of concept numbers, since the failure to pair a_2 with an equivalent descriptor in B may simply be due to the fact that a_2 was incorrectly numbered. Any error of this type should be immediately corrected.

- If a numbering error is found and corrected, and the correction results in a one-to-one correspondence or one-many correspondence between the languages, re-define the case at hand as being

of type (a) or (c), as appropriate, and proceed accordingly. (The procedures for case (a) have already been mentioned - see above; those for case (c) are listed below.)

- If a numbering error is found and corrected, but the correction leaves the case unchanged, proceed to 2.

- If no error is found, proceed to 2.

a 2. Apply the procedures described in 3.1.4 for establishing an equivalence between a_2 and a descriptor in B.

b) <u>One-many correspondences - Case (c)</u>

b 1. Check for errors in the allocation of concept numbers. Any error of this type should be immediately corrected.

- If a numbering error is found and corrected, and the correction results in a one-to-one correspondence or one-to-none correspondence between the languages, re-define the case at hand as being of type (a) or (b), as appropriate, and proceed accordingly.

- If a numbering error is found and corrected, but the correction leaves the case unchanged, proceed to 2.

- If no error is found, proceed to 2.

b 2. Determine whether all or some of the various descriptors in B which are shown as corresponding to a_2 are synonyms. If synonymous descriptors are found in B, they and their associated equivalence classes should be merged into a single descriptor representing a single equivalence class.

- If all of the descriptors in B which are listed as corresponding to a_2 are found to be synonyms, then the merging of synonyms in B will lead to the establishment of a one-to-one correspondence between a_2 and a single descriptor in B. The descriptors in the two languages should then be treated in the manner suggested for case (a), i.e. they should be inspected to ensure that they are genuinely identical in meaning, and only if this is confirmed should they be entered in the thesaurus as equivalents.

- If only some of the descriptors in B which are listed as corresponding to a_2 are found to be synonyms, so that the merging of synonyms in B leaves the situation as one in which a_2 is linked to two or more terms in B, proceed to 3.

- If there are no synonyms among the descriptors in B which are listed as corresponding to a_2, proceed to 3.

b 3. Identify the degree of equivalence (exact, inexact, non-equivalence) between a_2 and each of the descriptors in B which are listed as corresponding to it.

- Any descriptor in B which is found to be totally non-equivalent to a_2 should be handled by applying the procedures described in 3.1.4 for establishing equivalences between descriptors in two languages. (It should be noted that, in this particular case, the notation used in 3.1.4 to refer to languages and descriptors is the reverse of that used here. Hence, each of the terms which are here referred to as 'descriptors in language B' should be identified with the descriptor which is called 'a_2' in 3.1.4.
Likewise, the language in which equivalents are sought - here called 'A' - corresponds with 'B' in 3.1.4.
The procedures listed in 3.1.4 are valid provided this notational change is observed.)

- Any descriptor in B which is found to be exactly equivalent in meaning to a_2 should be recorded as such in the thesaurus. (There can be, at most, only one term which meets this description, given that any synonymous descriptors in B were removed at Step 2.)

- In the case of any descriptor in B which is found to be partially equivalent in meaning to a_2, the methods for reconciling the inexact equivalence between this term and a_2 should be identified by reference to the classification of inexact equivalences given in 3.1.2.

- If any descriptor in B which was originally listed as corresponding to a_2 was found to be either exactly or inexactly equivalent to a_2, the procedures described under Step 3 will be sufficient to ensure that each of B's descriptors has an equivalent in A, and, in addition, that a_2 has an equivalent in B. In this case, no further action is required once Step 3 has been completed.

- If all the descriptors in B which were originally listed as corresponding to a_2 proved to be non-equivalent to a_2, proceed to Step 4, since, in this case, the procedures described under Step 3 are not in themselves sufficient to provide a_2 with an equivalent in B.

b 4. Apply the procedures described in 3.1.4 for establishing an equivalence between a2 and an appropriate descriptor in B. (It may be noted that, on this occasion, the notation used for languages and descriptors in 3.1.4 is identical to that used here.)
If the correspondence between descriptors, as shown in the computer-generated list, were represented diagrammatically, they would often take the form of a complex network, e.g.

English	German
Rivers	Strom
	Fluß
Streams	Bach

Even in cases of this kind the procedures previously described are applicable, provided the connections within the network are considered concept by concept, and are analysed according to the five types of correspondence between descriptors listed at the beginning of this section. For instance, the connections in the network shown above can be reduced to two one-many correspondences between descriptors, i.e.

Concept number	English	German
123	Rivers	Strom, Fluß
124	Streams	Bach, Strom

Once the equivalences established in the first draft of the multilingual thesaurus have been checked and any necessary modifications made, a second draft should be prepared. The control procedures should then be repeated to ensure that the amendments made to the first draft did not themselves create new equivalence problems.

Many of the definitions and scope notes used within a multilingual thesaurus apply in more than one language, and many contain words which also occur as descriptors. It is important to ensure that if a definition or scope note contains a descriptor, an equivalent descriptor is used when the same definition or scope note appears in another language. This can most easily be checked if a KWIC or KWOC index is prepared to the definitions and scope notes in each language. Inconsistencies in terminology can then be traced in two ways: (i) by comparing the index entries in the different languages; (ii) by comparing the entry terms in the index for each language with that language's descriptors.

5.4 Testing

After the control procedures outlined above have been completed, and before publication, the multilingual thesaurus should be subjected to an indexing and retrieval test. This is particularly important in the case of thesauri constructed by the Gestalt method.

For the indexing test a sample of documents in a subject area covered by the thesaurus is required. The sample should be indexed independently in each of the languages represented in the thesaurus. Any differences in the resulting sets of indexing should be analysed, and attributed either to variations in subject analysis between indexers, or to the fact that indexers working in different languages place different interpretations on descriptors which are treated as equivalent in the thesaurus. Only the latter case is considered here. Consistency in the interpretation of terms can be improved in various ways: by modifying the structure of the thesaurus, by adding scope notes, or by adding references to descriptors from additional non-descriptors.

For the retrieval test a sample of queries is required. These should be translated independently into each of the languages represented in the thesaurus. A comparison should be made of the sets of documents retrieved in response to the searchers made in the different languages. Differences between the retrieved sets should be investigated and attributed either to variations between searchers, or to inconsistencies arising from defects in the thesaurus. If necessary, the procedures previously suggested for the improvement of consistency in the choice of terms should be adopted. Once modifications resulting from the indexing and retrieval tests have been applied, the multilingual thesaurus can be assumed to be ready for publication and usage.

6 Organizational aspects

In this section, the institutions which co-operate in the construction of a multilingual thesaurus will be called 'agencies', irrespective of whether they are private, government-owned, profit-making, or non-profit-making. Collectively, the cooperating agencies will be said to form an 'organization'.

The decision making procedures within the organization can have a profound effect on the thesaurus. Three types of organizational structure can be distinguished:

- the centralised structure - all decisions are made by a single agency. This allows decisions to be taken quickly, but there is the danger that they may not sufficiently take account of the needs of all thesaurus users. An automatic structure seems most appropriate where the multilingual thesaurus is to have only one main language, which

is also the working language of only one agency. This agency may then be given sole responsibility for constructing and updating the core of the multilingual thesaurus.

- the totally decentralised structure - all agencies communicate with each other to discuss solutions to the problems encountered in the construction of the multilingual thesaurus. This type of organizational structure has proved to be so ineffective as not to warrant further discussion here.

- the semi-centralised structure - a central committee is appointed, consisting of delegates from the co-operating agencies. The interests of each of the languages to be covered by the thesaurus should be represented by at least one delegate. The central committee should preferably be established on a full-time basis to organise such matters as co-operation between agencies, the allocation of work, and the co-ordination of all technical studies submitted by the agencies and subject specialists. It should also have the final decision in all matters affecting the multilingual thesaurus. What part of the technical work should be undertaken by the central committee itself depends upon the method of thesaurus construction used. If the analytical method is adopted, many decisions will have to be referred to the central committee. If, on the other hand, the method of construction is to translate hierarchies from one language to another, this task can be delegated, although the committee should still maintain overall control of the work.

Whatever type of organizational structure is chosen it should be valid not only for the construction of the multilingual thesaurus but also for such periodic updating and verification procedures as are involved in maintaining it in an up-to-date form.